Machine vision: An aid in reverse Turing test

Article *in* AI & SOCIETY · February 2011

DOI: 10.1007/s00146-009-0231-4 · Source: DBLP

2 authors:

Santosh Putchala
Philips

Nikhil Agarwal
University of Texas at Austin

AI & Soc (2011) 26:95–101
DOI 10.1007/s00146-009-0231-4

OPEN FORUM

Machine vision: an aid in reverse Turing test

Santosh Putchala · Nikhil Agarwal

Received: 22 September 2009 / Accepted: 3 October 2009 / Published online: 8 November 2009

Abstract Information security is perceived as an important and vital aspect for the survival of any business. Preserving user identity and limiting the access of web resources only to the humans and restricting 'bots' is an ever challenging area of study. With the increase in computing power and development of newer approaches towards circumvention and reverse-engineering, the recognition gap present between the machines and the humans is said to be decreasing. Turing test and its modified versions are in place to deal with such problems and ways to resolve them by developing complex algorithms for bot prevention systems like CAPTCHA (Completely Automated Public Turing test to tell Computers and Humans Apart). This paper will deal with the use of "Machine Vision" for judging the ability of the machines to compete with humans in breaking sequences of security systems like CAPTCHA. Reverse Turing test will be put to practise here. Complex image recognition technologies and novel approaches towards using Human interactive proofs (HIP) are discussed. The progress of Turing test over the past 60 years has been paid due attention at the end. After all this experimentation, it can be said that the current machine vision is quite poor and is far worse than it is expected to be.

S. Putchala (✉) · N. Agarwal
Europe Asia Business School, Pune, India
e-mail: spkris@yahoo.com

N. Agarwal
e-mail: nikhil.agarwal@eabs.ac.in

1 Introduction

Over the past 15 years, there has been a phenomenal increase in the number of websites created and their relative web pages. With such a trend, the concerns about online security and the 'bot[1]'prevention heightened with time. Development of new and robust security protocols that prevent automated intrusion has been pragmatic in the industry. Human interaction proof (HIP) is needed in order to ensure that no means of AI is used to get a legitimate access to the resources intended to be used only by humans. With all due respect to Alan Turing's legendary paper *'computing machinery and intelligence (1950)'* (Turing 1950), several attempts were made to introduce the variations of Turing's test that most often bank on the gaps between the human understanding and the machine intelligence.

Taking this into consideration, we propose here an experimental study to describe the shortcomings of the highly ranked modern day optical character recognition (OCR) algorithms in reading and interpreting CAPTCHAs. The character recognition ability gap existing between humans and the AI is discussed in detail here. In this new age of Web 2.0, most businesses are done online, and many other forms of business are related to the web in one way or the other. Enhancing the security and prevention of undesirable access to these websites is highly desirable. This paper will enumerate the existing security flaws and suggest suitable remedies. Techniques of attacking the popular CAPTCHAs and the proposed enhancements are also discussed.

[1] Bots are software applications that run automated tasks over the Internet at a much higher rate than would be possible for a human alone.

1.1 Turing's test and its modifications

Turing (1950) discusses about *'imitation game'* and proposed a question "Can machines think?" In the set-up, a human judge asks questions to a human player and a machine. Physical separation exists between judge and the two subjects are also physically isolated. The aim of the judge is to determine who is who. Failing to do so correctly would suggest the existence of artificial intelligence in the machine. Several attempts have been made by the computing community to make a machine pass the test. But this was achieved with a little success.

Usage of modified Turing's test has been widespread in the web space with the introduction of CAPTCHA (Completely Automated Public Turing test to tell Computers and Humans Apart). This is the challenge response test used to ensure that the response is not generated by a non-human. The evolution of enhanced user interface from text-based services has opened the path for the development of image-based CAPTCHA implementation. JPEG, GIF and PNG[2] are the most popular formats that have been rendered. Approaches like preventing the bot entry are being taken up to ensure a bot-free online space. The details (http://www.robotstxt.org/robotstxt.html) indicate a clear instruction for the bots to not to enter the website. This is essential as spamming is growing exponentially. Harvesting of e-mail addresses and the undesired access by the bots is being prevented by this method.

Popular e-mail account providers (Yahoo!, Hotmail, Google, etc.) honour these conventions. There are many other websites that do not honour these standard conventions and therefore abused about privacy and access control. The reverse Turing test is practiced as a standard by AI researchers world over.

1.2 History of impersonation by machines

Lillibridge et al. (2001) at Digital Equipment Corporation in Palo Alto, CA proposed a scheme to develop an image carrying text that can be easily interpreted by humans and is hard to decipher by machines. The emphasis was laid on OCR (Machine vision) capabilities of machines. In the year 2002, Baird, Popat and Broder reported that the usage of this system helped in reduction in spam addition of URLs by a record 95%. Even before that, Broder the person credited first for the development of CAPTCHA (while working at AltaVista) challenged Prof. Manuel Blum of The School for Computer Science at Carnegie-Mellon University (CMU) to design an efficient mechanism to restrict the usage of Yahoo!'s chat rooms to humans.

The CMU's team responded with the development of *'GIMPY'*.[3] This program rendered alpha-numeric text into an image. The seeds were picked at random. Features such as overlapping, warping, and shape deformations were present in this. In due course, '*EZ-GIMPY*' was used to resolve the chat-room problem. Von Ahn et al. (2000) described many parameters for a scheme to qualify for a CAPTCHA. These include

(a) The test shall be capable of rejecting ALL machines.
(b) The test input has to be automatically generated.
(c) Tests should be quick enough to be solvable by humans.
(d) The age factors are not taken into consideration as the test is assumed to be fairly simple for humans of varying age groups.
(e) Longevity of the test has to be ensured even after the algorithms or the source code has been divulged.

These parameters provide an opportunity for betterment relying on the well-known fact that recognition gap exists between the humans and the machines. As methods like dictionary attack are possible only when the OCR (machine vision) picks even the faint text patterns. Changes in the quality of plain text images by various image manipulation tools that purposefully generate distorted/poor quality imagery are selected. Care was taken that these images are perfectly readable by humans. The methodology of the experiment design revolves around the premise that *reverse Turing test* can be applied in this context.

1.3 Design of the test (reverse Turing test)

The Turing test is conceived as having a human judge and a computer subject who attempts to appear human. It is assumed that the human subject will always be judged as a human, and it is the computer subject who has to prove himself i.e. he has to pass the Turing test too along with the human subject. Here, the 'Reverse Turing test' is administered by a computer to determine whether the subject is human or not.

The test scenario can be explained as follows. A machine (bot) tries to access a website and wants to prove itself as a human. The authentication system present in the website presents the bot with a randomly generated CAPTCHA image and the bot should enter the correct

[2] Joint photographic experts group (JPEG), Graphics interchange format (GIF) and Portable network graphics (PNG) are digital image formats.

[3] GIMPY is a reliable system. It was originally built for Yahoo! to keep bots out of their chat rooms, to prevent scripts from obtaining an excessive number of their e-mail addresses, and to prevent computer programs from publishing classified ads. It is not an acronym.

combination of alpha-numeric text so as to gain access to the website. There is slight difference between the actual Turing test and the test conducted here. The aim of the test is to make the clear distinction between human and a bot but not to fail in the identification process. Another difference is that this test is fairly simple than the complex challenges thrown at the subjects in the imitation game.

Problems can arise when the bots attempt to keep track and record of the challenges thrown at them earlier. This can be circumvented by generating the CAPTCHA imagery using random numbers/alphabets. As no such thing as random exists (as we provide a seed to the algorithm to produce an output), referring them to be Pseudo-random might be apt.

2 Experiments: methodology

The design of the experiment begins with the identification of the software program(s) that develop(s) the human-legible and machine-illegible imagery. This was taken up as the first step. Two softwares were identified in this process.

Pixopedia 24 (http://www.sigmapi-design.com/cms/index.php?page=pixopedia-24)

GIMP 2.6.7 (http://www.gimp.org/downloads/)

These software packages have either a freeware license or a GNU General Public License as published by the Free Software Foundation.

The primary source of the text images is a word-processing application of whose screen shots were taken and then were subjected to various image manipulation features bundled in the above stated softwares. The letters that form words were selected in a pseudo-random method. The main idea of the above mentioned work is to develop a huge workable sample of the CAPTCHA imagery that closely resembles the widely used CAPTCHA images. While designing the imagery, it was kept in mind that it should be easy for a human to read out (decipher) the CAPTCHA, and it has to be hard to the non-human AI (Bots, spiders, OCR, etc.). Difficult to read image distortions of almost all available types (*41 types in total*) are included in the test set.

The second step was the identification of machine subjects. For this, the authors relied upon the optical character recognition (OCR) softwares to take the role of machine vision. Three OCR softwares were selected, and the selection is based on the fact that these softwares occupied top three slots consistently over years in the OCR industry. This ranking is done by various PC magazines and independent testing labs. The softwares are as follows:

(1) ABBYY Fine Reader 8.0 Professional Try & Buy version
(2) Read IRIS 12 Pro demo version
(3) Omni page professional 16 Trial version

The optical character recognition (OCR) softwares are quite expensive, and so here the demo versions of the softwares are being used. The demo versions of the selected softwares have full functionality but for a limited period. So for the lab test run, their functionality is as good as the fully licensed versions.

2.1 Design of the experiment

A set of 25 ALPHA and 25 ALHPA-NUMERIC RAW images (50 in total) were generated using the method described in the previous paragraphs. These were all 5-character long. All the sequences included English and base 10 numbers. Out of the 41 types as said earlier, 29 types of manipulations were promising and so were taken into the final experiment set. Generated RAW images were subjected to these 29 types of image manipulations.

These 29 types of manipulations were chosen, keeping in mind, the difficulties faced by the OCR machines and the pertaining software. Belief list of the degradation/distortion techniques[4] is as follows

(a) Noise (Wood rings, Skin, Marble, Hurl, Pick)
(b) Composed filters (Mini-max)
(c) Draw (cADD)
(d) Blur (Gaussian blur, Motion blur, Pixelize)
(e) Distortions (Blinds, Emboss, Engrave, IWarp, Ripple, Shift, Waves, Whirl & pinch, Wind)
(f) Edge detect (Neon, Convolution matrix)
(g) Artistic (Canvas, GIMP passionist, Oilify, Predator, Weave)
(h) Map (Displace, Illusion)

Calibri type Face was used, and the font size was 72 point. In total, the sample consisted of 1,450 b/w CAPTCHAs ([29 types $\times$ 25 images] $\times$ 2) + 10 colour CAPTCHAs. Inclusion of factor 2 is due to the usage of two sets (ALPHA & ALPHA-NUMBERIC). Distortions and the background noise are the two features that were studied extensively here. A limited sample of 10 colour images was also subjected to the same test to ensure wide range.

[4] The reduction in the inherent optimum detail and reduction in quality caused by unavoidable factors not associated with the sensor system lead to image degradation. Image distortion is the alteration of the original shape (or other characteristic) of detail in the image or the entire image.

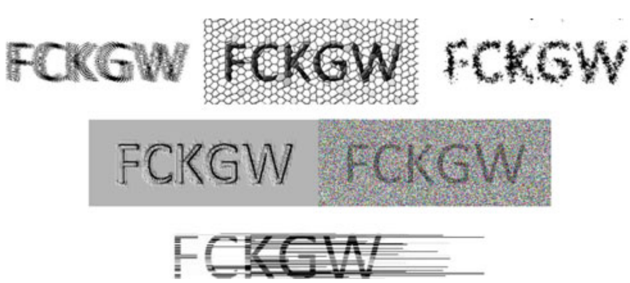
Some degree of noise is always present in any electronic device that transmits or receives a 'signal'.[5] Even though noise is unavoidable, it can become so small, relative to the signal that it appears to be non-existent. The signal to noise ratio (SNR) is a useful and universal way of comparing the relative amounts of signal and noise for any electronic system. High ratios will have very little visible noise, whereas the opposite is true for low ratios. Some images with sufficiently high SNR show clearly separated image information from background noise. A low SNR would produce an image where the 'signal' and noise are more comparable and thus harder to differentiate from one another [See second row of sample imagery (Sect. 2.2) from L to R]. Three common types of noise exist. Namely random noise, "fixed pattern" noise, and banding noise.

(a) Random noise is characterized by intensity and colour fluctuations above and below the actual image intensity.
(b) Fixed pattern noise includes what are called 'hot pixels', which are defined as such when a pixel's intensity far surpasses that of the ambient random noise fluctuations.
(c) Banding noise is highly device dependent, and it is the noise that is introduced when OCR sensor reads data from the digital feed. Banding noise is most visible when an image has been overly brightened.

Random noise is usually much more difficult to remove without degrading the image.

Computers face difficulty in judging random noise from fine texture patterns, so random noise can help us in generating hard to crack CAPTCHAs.

2.2 Sample imagery (generated)



2.3 Experimental results

Of the total 1,460 images used, all the images were human legible. But all the three renowned OCR systems failed miserably on almost every sequence. The Table 1 shows the results.

[5] Set of human information or machine data can be considered a signal.

Table 1 Results from the OCR machine

Type face/font	ABBYY reader	Read IRIS	Omni page	Total images
Calibri/72	0.15%	0.00%	0.00%	1,460

Here, OCR system accuracy is measured in the terms of % recognition in comparison to the original text.

As observed from the data obtained from the experiment, around 2 images were read and interpreted correctly by the OCR system. As the CAPTCHA images used in the experiment were generated by a human being using a software package, rate of error can be taken into consideration which might have allowed the OCR system to interpret those 2 sequences. Coming to the remaining text images, no OCR system guessed at least a single alphabet/number in the string of 5 characters. This suggests tragic failure of several machine vision approaches trying to impersonate humans. Almost all of the CAPTCHAs used in this experiment are so simple for humans that even a 7-year old kid can read them (provided he/she received the formal education).

From this, a generalization can be drawn that, not only these three OCR systems but, almost every other modern OCR system cannot handle the image distortions that have been employed here in the experiment. The reason behind drawing a generalization is that if the performance of top 3 OCR systems is this abysmal, what will be the ability of other OCR systems that are down the list in ranking.

The CAPTCHA imagery developed for this experiment has not needed much technical expertise. Few trial and error efforts lead to the generation of desirable CAPTCHAs as input to experiment. This indicates that the range and diversity of introduction of distortions is fairly large and even simple manipulations can perplex the machines vision.

3 Securing the web: complex image recognition

The web services are being offered to increase the ease with which we work. Over the past 10 years, web enabled services have generated much more revenue than any other service. The ease of use has made the web environment promising for further development. This popularity resulted in exploitation of the web services by bots. New security protocols were enforced so as to defend the web services and enable them to differentiate between humans and machines.

In September 2000, Udi Manber of Yahoo! described the famous" chat-room problem" Baird and Popat (2002) to researchers at Carnige Mellon University. The problem

description included the issue about ‘bots’ joining online chat rooms. This resulted in unnecessary usage of computing resources alongside the disruption of a peaceful chat environment because these bots pointed users to advertising sites. This was resolved by using the GIMPY CAPTCHA. The intrusion of bots has been related to the popularity of the instant messenger service. Recently, Skype users started reporting bot intrusions and spamming.

Techniques like *‘Pessimal print’* (uses a faint or degraded image CAPTCHA) (Baird et al. 2005), *‘Baffle text’* (uses a non-sense word that usually cannot be a victim of dictionary attack) (Lillibridge et al. 2001) and *‘Scatter type’* (drifting text which has been pseudo-randomly cut) (Baird et al. 2005) were introduced to counter the advances in reverse-engineering attempts.

CAPTCHA deployment is mainly through the generation of static text or an object. This has served the purpose well for a long time. But advances in reverse-engineering and *Laundry attack* attempts have led to the search for alternate methods to deploy CAPTCHA. Redirection of the challenge to another website (most commonly a file hosting or adult site) is a common practice. To prevent such attempts, animated CAPTCHAs have been put to practise. This animated CAPTCHA will be relying on mouse clicks of dynamically changing objects. This, when redirected to a new website for user input will render another graphic that is totally different. As the location of mouse pointer is tracked by x & y co-ordinates, this will make the attempt much difficult when compared to normal image CAPTCHA.

The source code can be hidden from recompilation and reverse-engineering, but this is not a fool proof mechanism. To a maximum possible extent, automated tools cannot do this process. This can help to avoid fast and automated reverse-engineering. This solves the purpose to a large extent (Antonatos 2006).

It was stated (Rice et al. 1996, 1999) that in the image recognition procedures, it is hard for the OCR system to interpret low quality images with the current level of available technology. The discussion till now is continuing based on the premise that the designed HIPs are hard to break, and this takes us into non-adversarial frame of reference. At this point of time, we shall discuss the new security issues that may arise due to over emphasis on present HIPs foolproof-ness. The pertaining questions are:

(a) Will the machines be able to guess the word just by seeing a pattern (as done by humans)?
(b) Can the complex CAPTCHA imagery developed with image degradations be reverse-engineered and brought back to its original form?
(c) Can the longevity of the test discussed by Von Ahn et al. (2000) hold good forever?

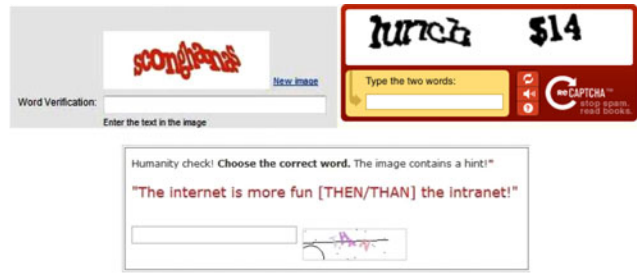


Fig. 1 Images are © 2009 reCAPTCHA

3.1 Difficult versus easy CAPTCHAs

If the string of letters in a CAPTCHA is a dictionary sequence, it brings both machines and humans to the same platform and with this, machines can guess words at a higher pace than humans can. But the usage of pseudo-random letters and numbers eliminates the possibility of guessing.

3.1.1 Difficult

The strings are not readily guessable (Fig. 1) and this ensures that the test is comparatively strong .

3.1.2 Easy

Presence of numbers makes the identification much easier for humans, and this is the same case with machines too (Fig. 2).

It should be noted that the words ‘easy’ and ‘difficult’ are not absolute in their meaning.

reCAPTCHA (Luis von Ahn et al. 2008, September 12) is being used to stop spam and to protect e-mail addresses from undesirable harness. reCAPTCHA codes can be generated and provided in place of e-mail addresses. Typically, this code presents a challenge as depicted above. The subject has to enter the correct sequence to get access to the e-mail address. The author’s e-mail address can be obtained from the URL: http://mailhide.recaptcha.net/d?k=01-jGq7ajNUDR3HmeDkF3nOg==&c=1ZAoWNdEGJ8Twv-1I_sYF4bENxpcDUSOf--YHHepRhQ=.

‘Kitten Auth’ (Fig. 3) used by Rapidshare.com was a form of complex image that used a CAPTCHA-containing alpha-numeric characters overlapping line images of either cat or a dog. Each character has line image of either a cat or a dog behind it. The user was asked to key in the characters that have the image of a cat. This gave the access to the file desired by the user. Complex forms of *‘Kitten Auth’* were also employed.

The possibility of dictionary attacks used to be quite high when ‘sense’ words are used. This flaw was overcome by the introduction of Text graphics character CAPTCHA (Chanathip and Matthew 2004). However, the growing

Fig. 2 Images are the © 2009 reCAPTCHA

sophistication of attackers and correspondingly increasing profit incentives have rendered most of the currently deployed HIPs vulnerable to attack (Chellapilla and Simard 2004; Mori and Malik 2003; Goodman and Rounthwaite 2004). There have been several attempts to introduce audio CAPTCHA and other new forms of HIPs. But these have succeeded to a lesser extent than the well known Text/Graphic CAPTCHA (Fig. 3).

The image shown to the left in the picture (Fig. 3) asked users to click on the all the 'kitten' in the picture and then click on the submit button to access the content.

4 Turing test: 60 years later

It has almost been 60 years since the legendary paper by Alan Turing was published. The Turing test and its modifications are used now prominently to deal with problems pertaining to AI. There have been a lot of expectations from Turing test and its later developments. Seeing the Turing test as the ultimate goal for any AI researcher makes the goal weak and less coherent. The case can be that there are no necessary and adequate proofs to show that machines can actually pass the Turing test. Passing Turing test can be *'impossible'* at this point of time but chances of it remaining the same are getting limited.

Many issues can be raised; these include but are not limited to

(a) What is the relationship between the language and the cognitive abilities?

(b) How computers can be made to properly read and understand human languages?

(c) Are the AI systems evolving as shown in fiction?

It can be said that passing the Turing test is not the only way to solve the AI vs. Human intelligence war. It can be seen as an aggregate of many machine-learning stages that will one day make the machines pass the Turing test.

5 Discussion

From the earlier mentioned experimentation and related analysis of the results, it can be hypothesized that not only these 3 optical character recognition systems but all modern OCR machines will not be able to cope up with the range of distortions introduced into the CAPTCHA imagery. A study of the history of image pattern recognition technology and of OCR technology suggests that the gap in ability between human and machine vision is wide and is slowly narrowing (Pavlidis 2000; Nagy 1996).

In the near future, attempts can be made to find a point at which humans are unable to read the distorted text, and thus this enhances HIPs' effectiveness.

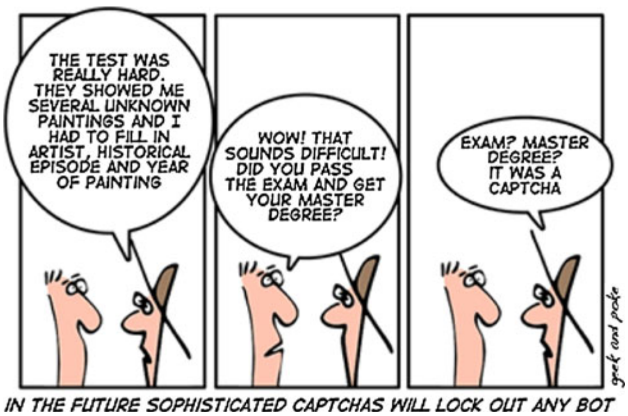
6 Conclusion

We have designed, built, and tested a reverse Turing test using "Machine vision"—that is, the optical character recognition ability. After all this experimentation, it can be

Fig. 3 Kitten Auth original (*left*) Kitten Auth used by Rapidshare.com (*right*)

said that the current machine vision is quite poor and is far worse than it is expected to be. The text in generated CAPTCHAs is legible enough even for a 7-year old (provided he/she received formal education). This answers the question "Are CAPTCHA-based HIPs reliable even now?"CAPTCHAs are till date the fastest and fully automatic methods for telling people and machine users apart over GUI interfaces.

Machine reading (According to Alan Turing)—which he planned to attack and which he expected to yield easily—has instead resisted solution for 60 years and now is poised to provide a technical solution and in providing a combative solution for Machine/Human inter-distinguish ability.



References

Antonatos EA (2006) Enhanced CAPTCHAs: using animation to tell humans and computers apart. Int Fed Inform Process 97–108

Baird HS, Popat K (2002) Human interactive proofs and document image analysis. In: Document analysis systems V-lecture notes in computer science (pp. 531–537). Springer, Berlin

Baird HS, Coates AL, Fateman RJ (2001) PessimalPrint: a reverse Turing test. In: Proceedings of the 6th international conference on document analysis and recognition, 1154–1158

Baird HS, Moll MA, Wang S-Y (2005) ScatterType: a legible but hard-to-segment CAPTCHA. In: Proceedings of the 2005 eighth international conference on document analysis and recognition (ICDAR'05). IEEE Computer Society

Chanathip N, Matthew D (2004) Mitigating dictionary attacks with text-graphics character CAPTCHAs. In: TENCON 2004 conference proceedings 2004

Chellapilla K, Simard P (2004) Using machine learning to break visual human interaction proofs (HIPs). Neural information processing systems (NIPS'2004). MIT Press, Cambridge

Goodman J, Rounthwaite R (2004). Stopping outgoing spam. In: Proceedings of the 5th ACM conference on electronic commerce. New York

Lillibridge MD, Adabi M, Bharat K, Broder A (2001) Patent No. US Patent 6,195,698. United States

Mori G, Malik J (2003) Recognizing objects in adversarial clutter: breaking a visual CAPTCHA. Proc Comp Vis Pattern Rec (CVPR) Conf IEEE Comput Soc 1:I-134–I-141

Nagy G (1996) Modern optical character recognition. The Froehlich/Kent Encycl Telecommun 11:473–531

Pavlidis T (2000) Thirty years at the pattern recognition front. King-Sun Fu prize lecture 11th ICPR. Barcelona, Spain

Rice SV, Jenkins FR, Nartker TA (1996) The fifth annual test of OCR accuracy. ISRI TR-96-01, Las Vegas

Rice SV, Nagy G, Nartker TA (1999) OCR: an illustrated guide to the frontier. Kluwer, Amsterdam

Turing A (1950) Computing machinery and intelligence. Mind 59(236):433–460

Von Ahn L, Blum M, Langford J (2000) Completely automatic public Turing test to tell computers and humans apart. The CAPTCHA Project, www.captcha.net, Dept. of Computer Science, Carnegie-Mellon Univ

Von Ahn L, Blum M, Langford J (2008) reCAPTCHA: human-based character recognition via web security measures. Science 321:1465–1486